COMMENTARY

From risk to fairness

Andreas Kappes[a,b], Guy Kahane[b], and M. J. Crockett[a,1]

Kadcyla is a drug that extends the life of breast cancer patients by an average of 6 mo. It also happens to be incredibly expensive. The United Kingdom's National Health Service sparked controversy when it refused to provide this drug to patients, citing its low cost effectiveness. Cases like this raise the question of how societies should make distributive decisions. Should we maximize utility or should we aim to improve the lives of the least fortunate, even if doing so is costly for everyone else? The influential philosopher John Rawls tackled this dilemma by framing fair distributive decisions as a kind of gamble (1). Rawls famously argued that we should choose the kind of society we would all prefer if our choice was made from behind a "veil of ignorance"—that is, under conditions of complete uncertainty about where we would end up. He held that people should make such choices by following a risk-averse "maximin" strategy of maximizing the minimum possible outcome for themselves and others. Echoing Rawls's theory, new research by Kameda et al. (2) links risk and fairness by showing that preferences about risk and about distribution may arise from common psychological and neural substrates.

In the experiments, participants made two kinds of decisions that at first glance seem rather different: distributive decisions involved allocating payoffs to anonymous others, whereas risky decisions involved choosing between lotteries affecting their own payoffs. However, most participants deployed similar strategies for both types of decisions. Those who followed a "Rawlsian" maximin strategy in their distributive decisions tended to adopt a similar strategy for risky decisions, avoiding lotteries with the lowest possible payoffs. Meanwhile, "utilitarian" participants maximized total welfare in their distributions and preferred lotteries that maximized expected payoffs. Participants' diverse decision patterns thus appear to mirror the philosophical debate between Rawls and his utilitarian critics around how best to address risk and uncertainty when making wealth distribution decisions (3).

A key insight of this work is that considering the worst possible outcome is a common feature of both risky and distributive choices. In the former, we imagine ourselves on the losing end of the bet; in the latter, we take the perspective of the least fortunate recipient. Thus, just as distributive choices involve a conflict of interests between individuals, risky choices can be described as a conflict of interest between lucky and unlucky selves. Others have used a similar logic to show that intertemporal choices between immediate and delayed rewards resemble a kind of social dilemma between present and future selves (4, 5). The common thread binding all these kinds of decisions is perspective taking, which has been linked to the right temporoparietal junction (RTPJ) (6). Accordingly, Kameda et al.

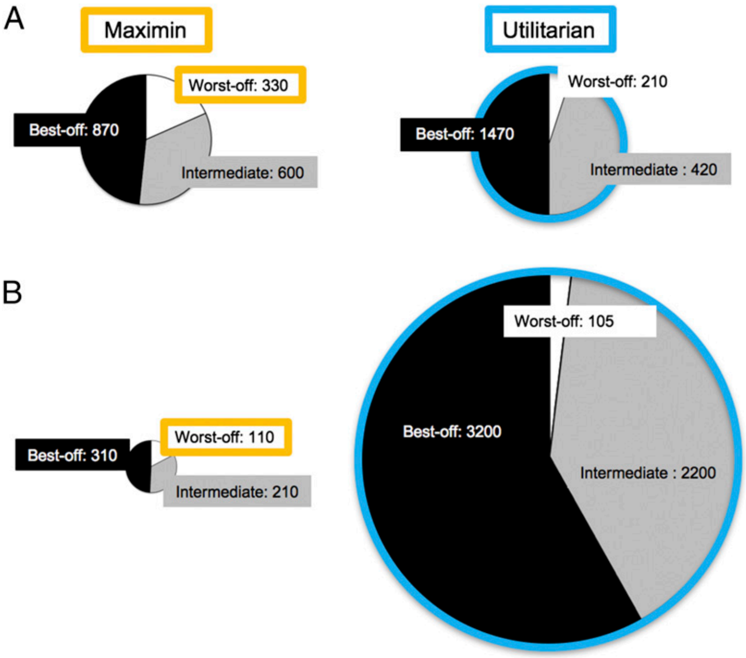

Fig. 1. Example wealth distribution scenarios, where numbers represent welfare in arbitrary units; the size of each pie corresponds to total group welfare. The Rawlsian maximin strategy aims to maximize the welfare of the worst-off person (highlighted in yellow), whereas the utilitarian strategy aims to maximize total group welfare (highlighted in blue). Maximin seems fairly intuitive when it involves only a small cost to group welfare (*A*). However, support for maximin may diminish as group welfare costs increase (*B*).

[a]Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom; and [b]Oxford Uehiro Centre for Practical Ethics, University of Oxford, Oxford OX1 1PT, United Kingdom

Author contributions: A.K., G.K., and M.J.C. wrote the paper.

The authors declare no conflict of interest.

See companion article 10.1073/pnas.1602641113.

[1]To whom correspondence should be addressed. Email: molly.crockett@psy.ox.ac.uk.

www.pnas.org/cgi/doi/10.1073/pnas.1614111113

report that RTPJ encoded the relative cost of the worst possible outcome during both risky and distributive choices. Connectivity analyses suggested that, during decision making, RTPJ communicates information about the worst possible outcome to the caudate, a region involved in computing choice values (7). These results are consistent with growing evidence that decision making for self and others involves shared neurocognitive mechanisms (8) and imply that alternative perspectives—those of others, or of future selves—are integrated into a common valuation circuitry during decision making. The study advances previous work by measuring risky and distributive choices in parallel, rather than in isolation. However, the methods used here do not permit the conclusion that maximin computations for risky and distributive choices are carried out by the same neurons. Further work using methods such as repetition suppression (9) or single-unit recordings (10) will be necessary to determine whether there are domain-general neurons in RTPJ that evaluate the worst possible outcome for both self and others.

Intriguingly, although most participants used similar strategies for distributive and risky decisions, they were significantly more likely to follow a maximin strategy for distributive decisions. This behavioral tendency to be more risk-averse for others than self was reflected in the neural data, where connectivity between RTPJ and caudate was stronger for distributive than risky choices. These findings may indicate that the worst possible outcome carries more weight when deciding for others than for self, perhaps because choosing for others has potential moral consequences whereas choosing for oneself does not. This idea finds additional support in a recent study where participants chose whether to inflict painful electric shocks on either themselves or an anonymous other person in exchange for money (11). Strikingly, most people were more averse to inflicting pain on others than themselves. One potential explanation is that people were more uncertain about how others would experience the shocks, and therefore were afraid that an amount of pain acceptable for them might be unbearable for the other person. This uncertainty could have induced a sort of risk premium on the moral costs of imposing potentially intolerable pain on another, leading people to deploy a pain-minimizing strategy when deciding about others' pain but a profit-maximizing strategy when deciding about their own pain—just as many participants in the study by Kameda et al. were loss-minimizing for distributive decisions affecting others but profit-maximizing for risky decisions affecting only themselves. Both studies are consistent with the "precautionary principle" in policy making, where actions that might cause harm are prohibited and decision makers must prove that actions are harmless (12). They also suggest that uncertainty about others' experiences might increase prosocial behavior by enhancing the salience of the worst possible outcome.

The overlap in mechanisms guiding risky and distributive choices may help explain why people make moral judgments about risky behavior, even when such behavior does not affect others (13). It also suggests we might gain insights into social decision making from the vast existing body of research on decision making under risk and uncertainty. If risk aversion leads people to adopt a maximin strategy when making distributive decisions, then factors affecting risk aversion might also impact distributive decisions. For instance, people perceive risks that evoke stronger emotional reactions as more dangerous than risks that do not (14). Accordingly, people might rely more on a maximin strategy when making distributive decisions that evoke strong emotional responses, such as when the decision involves identifiable individuals rather than anonymous groups (15), or when the decisions involve children rather than adults.

Kameda et al. report that a substantial portion of participants deployed a maximin strategy aimed at avoiding the worst possible outcome, especially when distributing wealth for others. However, before we can conclude whether people are truly Rawlsian, we need to consider a wider range of choice scenarios than those presented in the experiments by Kameda et al., where Rawlsian vs. utilitarian strategies resulted in relatively constrained welfare differences (Fig. 1*A*). One can imagine alternative scenarios in which the maximin strategy only barely improves the welfare of the worst-off, while dramatically reducing the total welfare of the group (Fig. 1*B*). Would people still prefer a maximin strategy in this case? Similar questions can be raised about the utilitarian strategy: would people still choose to maximize aggregate welfare even if this results in extremely bad outcomes for the least fortunate? When faced with cases of this sort, participants might rather exhibit a pattern of choice closer to the more nuanced distributive view known as "prioritarianism" (16), according to which overall outcomes should be maximized (a utilitarian idea) but where the weight assigned to each potential benefit also reflects how much worse off its beneficiary is (a Rawlsian idea). Uncovering the boundary conditions for distributive preferences—be they Rawlsian, utilitarian, prioritarian, or something else entirely—is an exciting avenue for future research.

New research by Kameda et al. links risk and fairness by showing that preferences about risk and about distribution may arise from common psychological and neural substrates.

Few topics divide us as sharply as the question of distributive justice. On issues like taxation, affirmative action, and healthcare provision, people are polarized into groups that reach vastly different conclusions about the best way to organize society. Likewise, most of the participants in the Kameda et al. studies divided into those who preferred to prioritize the needs of the worst off, and those who preferred to maximize overall welfare. It is hard to resist speculating that divisions between different political or philosophical camps might be, at least in part, rooted in diversity at the neurobiological level—which may explain the seeming intractability of many disputes about distributive justice. Indeed, critics of Rawls have similarly argued against Rawls's assumption that people would follow his maximin rule in conditions of uncertainty, suggesting that individuals might respond to risk with diverse strategies (3). However, the existence of such diversity in views might also suggest that different strategies—maximin and utilitarian—serve different functions within a group, and hence, this diversity may also benefit society. Just as we and other animals have to strike an optimal balance between risk seeking and risk aversion, societies have to balance a need to maximize overall welfare with taking care of the worst-off. Moreover, our ideological differences may not run as deep as they appear. Kameda et al. found that Rawlsians and utilitarians alike paid the most attention to the worst possible outcome when making their choices, and both groups showed RTPJ activation that correlated with concern for the worst-case scenario. Thus, it does not seem to be the case that utilitarians simply ignore the plight of the worst-off; rather, they appear to use this information differently than Rawlsians.

As the authors note in their discussion, Hume's famous dictum that "is" does not imply "ought" means we cannot draw direct normative prescriptions from their findings. However, it is hard to avoid thinking about the normative implications of the research. Rawls himself insisted that political philosophy must describe political arrangements that can gain support from real people (17). If, in line with Rawls, we hold that distributive justice must reflect a kind of hypothetical contract that we all could rationally agree to, then the diversity in preferences observed by Kameda et al. can raise doubts about the possibility of such a consensus. At the same time, their findings also show there is a natural inclination to heed the worst-case scenario, suggesting a common starting point—indeed, one that may point in a Rawlsian direction.

1 Rawls J (1999) *A Theory of Justice* (Harvard Univ Press, Cambridge, MA), Rev. Ed.

2 Kameda T, et al. (2016) Maximin rule as a common cognitive anchor in distributive justice and risky decisions: Rawls in our minds. *Proc Natl Acad Sci USA*, 10.1073/pnas.1602641113.

3 Harsanyi JC (1975) Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory. *Am Polit Sci Rev* 69(2):594–606.

4 Hershfield HE, et al. (2011) Increasing saving behavior through age-progressed renderings of the future self. *J Mark Res* 48(SPL):S23–S37.

5 Mitchell JP, Schirmer J, Ames DL, Gilbert DT (2011) Medial prefrontal cortex predicts intertemporal choice. *J Cogn Neurosci* 23(4):857–866.

6 Van Overwalle F (2009) Social cognition and the brain: A meta-analysis. *Hum Brain Mapp* 30(3):829–858.

7 Hsu M, Anen C, Quartz SR (2008) The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science* 320(5879):1092–1095.

8 Ruff CC, Fehr E (2014) The neurobiology of rewards and values in social decision making. *Nat Rev Neurosci* 15(8):549–562.

9 Barron HC, Garvert MM, Behrens TEJ (2016) Repetition suppression: A means to index neural representations using BOLD? *Philos Trans R Soc Lond B Biol Sci* 371(1705):20150355.

10 Chang SWC, Gariépy J-F, Platt ML (2013) Neuronal reference frames for social decisions in primate frontal cortex. *Nat Neurosci* 16(2):243–250.

11 Crockett MJ, Kurth-Nelson Z, Siegel JZ, Dayan P, Dolan RJ (2014) Harm to others outweighs harm to self in moral decision making. *Proc Natl Acad Sci USA* 111(48):17320–17325.

12 Sunstein CR (2005) *Laws of Fear: Beyond the Precautionary Principle* (Cambridge Univ Press, Cambridge, UK).

13 Cushman F (2008) Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108(2):353–380.

14 Slovic P, Finucane ML, Peters E, MacGregor DG (2004) Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Anal* 24(2):311–322.

15 Small DA, Loewenstein G (2003) Helping a victim or helping the victim: Altruism and identifiability. *J Risk Uncertain* 26(1):5–16.

16 Parfit D (1997) Equality and priority. *Ratio* 10(3):202–221.

17 Rawls J (2001) *Justice as Fairness: A Restatement* (Harvard Univ Press, Cambridge, MA).